DOCUMENT RESUME

ED 433 360 TM 030 049

AUTHOR Klockars, Alan J.; Potter, Nina Salcedo
TITLE Effects of Mild Curvature on ANCOVA and Randomized Blocks.
PUB DATE 1999-04-00
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Covariance; Interaction; *Research Design; Simulation
IDENTIFIERS Additive Models; *Blocks; Curvilinear Functions; *Randomization; Type I Errors

ABSTRACT

The type I error control and power of a number of analysis of covariance (ANCOVA) and randomized block (RB) designs with curvilinear data were studied for tests of the additive treatment effect and interaction. For tests of additive effects, the analysis was also conducted using systematic assignment to treatments and using random assignment with a higher order covariate. Each of the analyses was conducted using the normal measure of random variability, score variation from parallel lines, and score variation from unique lines. A FORTRAN program was written to simulate data that could be analyzed with each alternative. The most interesting finding is the monumental inflation of the Type I error rate for the standard ANCOVA with random assignment of subjects when used to detect differences in slope. The source of this inflated error rate lies in the assumptions of ANCOVA. Because of this inflated error rate, it is recommended that systematic assignment be used when an individual difference variable is built into an experimental design. There may be situations in which use of a higher order covariate or RB designs will serve the experimenter better, but ANCOVA with systematic assignment and errors about unique regression lines appears to be the current best practice. (Contains 3 tables, 1 figure, and 13 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 433 360

Effects of Mild Curvature on ANCOVA and Randomized Blocks

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Alan Klockars

Alan J. Klockars
Nina Salcedo Potter

TM030049

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in Montreal Canada
April 1999

BEST COPY AVAILABLE

Effects of Mild Curvature on ANCOVA and Randomized Blocks

Alan J. Klockars and Nina Salcedo Potter

University of Washington

An analysis that includes an individual difference variable or covariate (X), such as an analysis of covariance (ANCOVA) and randomized block designs (RB) can increase the statistical power over a completely randomized analysis of variance (ANOVA) to detect additive treatment effects (heights or adjusted means) and provide information about the interaction between the covariate and the treatment variables. A recent study by Klockars, Potter, and Beretvas (1999) used a Monte Carlo design to simulate the Type I error control and power of ANCOVA and RB when used to detect additive effects with the assumptions met to the extent possible in a simulation. ANCOVA had consistently greater power when the correlation between X and the outcome measure (Y) was greater than .2. The number of blocks used with the RB design influenced the power. As the correlation increased power was maximized with a larger number of blocks. In general the results of Klockars, Potter, and Beretvas were in agreement with an early paper by Feldt (1958) except that the superiority of ANCOVA was detected at a lower correlation than found by Feldt.

ANCOVA and RB can also be used to explore differential effects of treatments on the X scores. With RB this differential effect is found in the Block x Treatment Interaction (BxT) while in ANCOVA it is discovered in the test of the homogeneity of slopes of the unique regression lines through each of the treatment groups. The possibility of differences in slopes (an interaction between B and T) also raises the issue of the impact of heterogeneous slopes on the test of additive treatment effects. Since homogeneity of slopes is an assumption of the test for heights within ANCOVA the concern is whether the test is sufficiently robust to deal with slope heterogeneity. A number of studies have been concerned with the ability of ANCOVA to detect additive effects when there are heterogeneous slopes. Analytical studies, particularly Rogosa (1980), recommended that

BEST COPY AVAILABLE

ANCOVA be redefined using the variability of scores about the unique regression lines rather than about parallel lines as a strategy for increasing the power of the test for adjusted means. Simulations studies including Levy (1980) and Hamilton (1976), found that the standard test for additive treatment effects with error variance about the parallel lines tended to be slightly conservative with heterogeneous slopes unless confounded with unequal sample sizes and other violations of assumptions. In simulations by Harwell and Serlin (1988) and Klockars & Beretvas (1998) the use of the unique lines suggested by Rogosa produced serious inflation of the Type I error rate.

Homogeneity of slopes is very much related to the choice of the appropriate error term in RB. If the variability of subjects about the block-treatment mean (S/BT) is to be used as the measure of random error either blocks must be assumed to be a fixed factor or the interaction must be zero. The alternative to S/BT is to use the BxT interaction as the error term which allows for either homogeneity or heterogeneity of slopes but dramatically reduces the power of the test. Klockars and Beretvas (1998) provided Monte Carlo evidence that even when there is considerable heterogeneity of slopes the Type I error rate for RB with S/BT as the error term stays within Bradley's (1978) commonly accepted limits for robustness.

Klockars and Beretvas (1998) also compared the power of ANCOVA and RB to detect differences in slopes. The optimal number of blocks when attempting to identify the presence of a BxT interaction was 2 in every simulation except one in which 3 was slightly more powerful. Even with the optimal number of blocks, ANCOVA's test for slopes was considerably more powerful for detecting the heterogeneity of slope. The BxT interaction can be partitioned into the portion which tests for differences in the linear components of the interaction. This strategy is closer in power to the ANCOVA test of slopes but it is unlikely that an experimenter would go directly to this test without first requiring an omnibus test of the BxT. Thus the power of the linear components becomes tied to the

power of the omnibus test. Additionally the linear component has a somewhat elevated Type I error rates that increase as the strength of the relationship increases.

Dalton and Overall (1977) proposed a systematic strategy for assigning subjects to treatment groups in an ANCOVA design which they called the "alternate-rank method". Using this method subjects are rank ordered on X and then systematically assigned to the treatments. In a two treatment experiment the subject with the highest X score is assigned to group A, the second highest to group B, the third to group B, and the fourth to group A. This pattern of ABBA is repeated with all subjects. With more than two treatments a simple serpentine pattern is used although more complex systematic patterns are possible which would more equitably distribute the subjects based on their X scores. Maxwell, Delaney, and Dill (1984) and McAweeney and Klockars (1998) found that systematic assignment was more powerful for detecting additive treatment effect than random assignment.

While linear relationships are assumed to be ubiquitous within psychology and education, there are a number of situations in which mild curvature may appear. Most common are ceiling or floor effects that may reflect either a measurement problem or the limits on the effectiveness of further treatment. These conditions may produce curvature that is difficult to detect from an inspection of a relatively small sample from that population. The primary source of information concerning analyses when curvature may be present comes from textbooks. RB designs are generally lauded as appropriate for data in which there may be curvature since no assumption is made about the shape of the interaction. At least three blocks would be required to map an interaction involving curvature. Analysis of covariance is more problematic as a linear relationship is assumed for both the test of differences in heights and slopes. Both early and recent textbooks, Li (1964) and Maxwell and Delaney (1990), recommend including a higher order variable as a second covariate when there may be a curvilinear relationship. This results in the formula for Y' becoming $Y'=a+b_1X+b_2X^2$. Adding a second covariate has several ramifications.

The fit of the model to the data should be improved when there is substantial curvature but it reduces the number of degrees of freedom available for estimating random variability. The test of the interaction becomes the combined interaction of treatments with either linear or quadratic portion of the covariates.

The current study investigates the Type I error control and power of a number of ANCOVA and RB designs with curvilinear data. These are addressed for both tests of the additive treatment effect and interaction. For the test of additive effects, in addition to a standard ANCOVA with random assignment, the analysis also was conducted using systematic assignment to treatments and, again using random assignment with a higher order (quadratic) covariate. Each of these three analyses was conducted using the normal measure of random variability, score variation from parallel lines, and Rogosa's suggested alteration, score variation from the unique lines. An RB design was conducted using the S/AB and then the BxT as the estimate of error variance. Three blocks were used as a compromise between the recommended number for a test of additive treatment effects and a test of the interaction. The interaction was tested using the same three variations of ANCOVA; random assignment, systematic assignment, and a higher order covariate. Both the omnibus BxT and the linear component of the RB design are reported.

Method

A FORTRAN program was written to simulate data that could be analyzed with each alternative. Y scores were generated as a weighted combination of two normally distributed random variables; the first to measure the covariate, X, and the second to introduce random variability. Different weights were used to produce correlations between X and Y of ρ=.3, .5, and .7. These correlations between X and Y are prior to the addition of any curvature. Curvature was created by including X^2 in the data generation algorithm. There were four levels of curvature determined by the weight used with the quadratic term. Weight denoted 0 represented no curvature with weights of .1, .2, .3 and .4 representing successively greater curvature. An indication of the amount of curvature can be gleaned

from Figure 1 where the distributions of 1000 scores generated by the most severe curvature for each level of ρ are presented.

Four treatment groups of 24 subjects each were drawn from the various populations generating formulas. Additive treatment effects were created by adding z scores of {.25, 0, 0, -.25} to the existing four groups of Y scores. Two levels of heterogeneous slopes were used in addition to the homogeneous set of slopes. The slopes used were: {1,1,1,1} for homogeneous slopes; {.8 .9 1.1, 1.2} for mild heterogeneity; and {.5, .8, 1.2, 1.5} for moderate heterogeneity.

Each total sample of 96 scores was analyzed for differences in adjusted means with random, systematic, and higher order ANCOVA using both the parallel and unique lines. The RB used 3 blocks and provided a test based on S/AB and BxT as random error. The test for interaction was conducted using random, systematic, and higher order ANCOVA. For the RB design the BxT and linear component of BxT were found. Each Type I and power estimate is based on 100,000 iterations.

Results & Discussion

Table 1 contains the Type I error rates of the test of additive treatment effects, or heights. The first three columns indicate the degree of correlation present, the curvature, and the degree of heterogeneity of slopes, respectively. Curvature of 1 and 3 as well as mild heterogeneity of slopes are omitted as all results showed patterns in agreement with the presented results. The next six columns present Type I error rates for Random, Systematic and Higher Order ANCOVA. Within each method two values are presented using either parallel lines or unique lines as the bases of random error. The last two columns present the results for RB. The column marked S/BT presents results when the subjects variability is used as error while the last column uses the Block x Treatment interaction as error.

Type I error rates are acceptable for all ANCOVA methods when the parallel regression lines define random error. When the unique lines are used only the

simultaneous usage of systematic assignment retains acceptable Type I error. With both Random and Higher Order ANCOVA the error rate with unique lines become unacceptably high when there is moderate to high correlations paired with heterogeneous slopes. The Type I error rate for RB with S/BT as error shows the slight increase in Type I error rate when there is heterogeneity of slopes. When the BxT is used as error the error rates become excessively conservative when there is systematic variation in the BxT interaction.

Table 2 presents the power results within the same format. Overall the amount of additive treatment effect is insufficient to produce what are normally considered acceptably power ratings. These were intentionally kept low to avoid ceiling effects as ρ increased. Previous research (see McAweeney & Klockars, 1998) shows that the same patterns are obtained with larger additive effects. The power values for methods that exceeded a Type I error rate of .065 are presented with strikethroughs. When the correlation between X and Y is .3 there is little difference between methods with the exception of RB using the BxT interaction as error. This method is consistently well below any other. Of the remaining methods the largest difference in power is less than 2%. Interesting, the higher order ANCOVA is not superior to the remaining when the maximum curvature is present. Instead the Systematic assignment when paired with the use of the unique regression lines is slightly superior in all conditions.

When the correlation is .7 a slightly different picture has emerged. As expected RB using S/BT as error has less power than the ANCOVA methods. This difference is exacerbated by the decision to use only three blocks. When ρ=.7 a greater number of blocks would make the difference between RB and ANCOVA smaller. When using the parallel regression lines, systematic assignment is moderately superior to random assignment. Substituting the unique lines with systematic assignment greatly increases the power when there is heterogeneity of slopes. The other ANCOVA methods also would have shown this effect but they failed to control Type I error. Lastly, when there is

maximal correlation and curvature the Higher Order ANCOVA was considerably more powerful than any other method.

Table 3 presents the Type I error rates and power estimates for detecting the differences in slope for the moderately heterogeneous set of slopes (3). The most startling results are the Type I error rates for ANCOVA with Random assignment. As curvature increases the Type I error rate increase so that with $\rho = .7$ and curvature 4 there is a 20% chance of a Type I error. Less noticeable is an increase in Type I error for the linear component of the BxT interaction in RB. The exaggerated Type I error rate of ANCOVA is eliminated by using systematic assignment of subjects or a Higher Order covariate. As expected the Type I error rate of the omnibus BxT interaction controlled Type I error. The power of ANCOVA with systematic assignment of subjects was superior to either the Higher Order ANCOVA or the RB. The difference is diminished as there was more curvature in the data.

Conclusions

The most interesting finding is the monumental inflation of the Type I error rate for the standard ANCOVA with random assignment of subjects when used to detect differences in slope. The source of this inflated error rate is the assumptions of ANCOVA. Specifically, the assumption that the X variable is fixed assumes that the question being tested is the difference in slopes for a fixed and finite set of X scores. In the simulation a random sample is generated for each iteration with whatever X values are found. Because the concentrations of X values for the various samples will randomly differ from one another, the portion of the curve through which the linear trends for the samples is fit will differ. The greater the curvature, the greater the difference in linear fits for sets of Xs with slightly different concentrations along the X dimension. Only if all samples had identical concentrations of X scores would the linear fit for each group be equivalent. When scores were systematically assigned to treatments the random difference in X scores is greatly reduced and the Type I error rate decreases to acceptable levels.

With the inflated Type I error rate of standard ANCOVA for testing differences in slopes and the general superiority of systematic assignment for tests of heights, we recommend that systematic assignment be used when an individual difference variable is built into an experimental design. With this equalization of groups with respect to the X variable it is also possible to redefine random variability within ANCOVA as the variability about the unique regression lines without producing the previously detected increase in Type I error rate. While there may exist situations in which the use of a higher order covariate or randomized block designs will serve the experimenter better, ANCOVA with systematic assignment and errors about unique regression lines appears to be the current best practice.

References

Bradley, J.C.(1978). Robustness? *British Journal of Mathematical and Statisitcal Psychology, 32*, 144-152.

Dalton, S. & Overall, J.E.(1977). Nonrandom assignment in ANCOVA: The alternative ranks design. *The Journal of Experimental Education. 46*, 58-62.

Feldt, L.S.(1958). A comparison of the precision of three experimental designs employing a concomitant variable. *Psychometrika, 23*, 335-353.

Hamilton, B.L.(1977). An empirical investigation of the effects of heterogeneous regression slopes in analysis of covariance. *Educational and Psychological Measurement, 37*, 701-712.

Harwell, M.R. & Serlin, R.C.(1988). An empirical study of a proposed test of nonparametric analysis of covariance. *Psychological Bulletin, 104*, 268-281.

Klockars, A.J. & Beretvas, S.N.(1998). Randomized blocks and analysis of covariance for identifying additive treatment effects and interactions with heterogeneous slopes. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego

Klockars, A.J., Potter, N.S., & Beretvas, S.N.(1999). Power to detect additive treatment effects with randomized block and analysis of covariance designs. *The Journal of Experimental Education, 67*, 190-191.

Levy, K.J.(1980). A Monte Carlo study of analysis of covariance under violations of assumptions of normality and equal regression slopes. *Educational and Psychological Measurement. 40*, 835-840.

Li. J.C.R.(1964). *Statistical Inference II*, Ann Arbor, Edwards Brothers.

Maxwell, S.E.& Delaney, H.D(1990). *Designing Experiments and Analyzing Data: A model comparison perspective.* Belmost, CA. Wadsworth.

Maxwell, S.E., Delaney, H.D.& Dill, C.A.(1984) Another look at ANCOVA versus blocking. *Psychological Bulletin, 95*, 136-147.

McAweeney,M.J. & Klockars, A.J.(1998). Maximizing power in skewed distributions. Analysis and assignment. *Psychological Methods, 3,* 117-122.

Rogosa, D.R.(1980). Comparing non-parallel regression lines. *Psychological Bulletin, 88,* 307-321.

Table 1: Type I error rates for test of heights

			Random		Systematic		Higher Order		RB	
Cor	Curve	b	Parallel	Unique	Parallel	Unique	Parallel	Unique	Sub	BxT
0.3	0	0	.050	.050	.051	.051	.050	.049	.051	.050
		3	.050	.052	.048	.051	.050	.052	.050	.036
	2	0	.051	.051	.049	.049	.051	.051	.049	.049
		3	.050	.053	.048	.050	.051	.053	.050	.034
	4	0	.050	.051	.046	.046	.049	.049	.050	.050
		3	.051	.054	.044	.047	.050	.053	.051	.036
0.5	0	0	.050	.050	.051	.051	.050	.049	.049	.049
		3	.051	.059	.045	.053	.050	.058	.050	.016
	2	0	.051	.052	.047	.046	.051	.051	.049	.050
		3	.050	.058	.041	.049	.051	.059	.051	.017
	4	0	.050	.052	.039	.039	.050	.050	.049	.048
		3	.051	.060	.034	.040	.050	.057	.052	.019
.7	0	0	.049	.049	.049	.050	.051	.051	.050	.049
		3	.051	**.075**	.037	.056	.051	**.073**	.055	.004
	2	0	.050	.051	.042	.042	.050	.050	.050	.048
		3	.051	**.073**	.031	.047	.050	**.072**	.056	.005
	4	0	.050	.054	.024	.024	.050	.050	.051	.046
		3	.051	**.073**	.019	.029	.051	**.073**	.055	.007

This table shows the type I error rates when the correlation between x and y is .3, .5 and .7. The values of curves represent how much curve was added to the relationship between x and y, see figure 1 to see how much curve is represented by the number 0 –4. The slope column represents the amount of heterogeneity of slopes is represented: 0 being no homogeneous slopes, 3 being the most heterogeneity added.

BEST COPY AVAILABLE

Table 2: Power estimates for test of heights

			Random		Systematic		Higher Order		RB	
Cor	b	Slope	Parallel	Unique	Parallel	Unique	Parallel	Unique	Sub	BxT
0.3	0	0	.288	.287	.288	.288	.283	.281	.278	.176
		3	.285	.292	.285	.293	.283	.289	.285	.134
	2	0	.284	.284	.295	.295	.283	.282	.281	.175
		3	.283	.290	.292	.300	.284	.290	.280	.134
	4	0	.281	.283	.294	.294	.284	.283	.278	.174
		3	.283	.291	.293	.300	.288	.294	.281	.136
0.5	0	0	.339	.338	.344	.343	.338	.337	.322	.198
		3	.336	.362	.335	.362	.336	.359	.334	.087
	2	0	.338	.339	.352	.352	.339	.339	.319	.195
		3	.331	.358	.340	.367	.338	.362	.329	.087
	4	0	.325	.330	.342	.342	.347	.346	.312	.190
		3	.321	.350	.331	.355	.344	.368	.322	.089
.7	0	0	.469	.469	.469	.469	.463	.462	.403	.241
		3	.445	~~.513~~	.430	.505	.445	~~.508~~	.420	.004
	2	0	.444	.448	.472	.471	.468	.467	.392	.235
		3	.427	~~.495~~	.436	.508	.449	~~.514~~	.417	.041
	4	0	.402	.415	.421	.419	.492	.490	.374	.227
		3	.389	~~.455~~	.391	.452	.470	~~.535~~	.393	.045

This table shows the power estimates when the correlation between x and y is .3, .5 and .7. The values of curves represent how much curve was added to the relationship between x and y, see figure 1 to see how much curve is represented by the number 0 –4. The slope column represents the amount of heterogeneity of slopes is represented: 0 being no homogeneous slopes, 3 being the most heterogeneity added. Power estimates that correspond to error rates above .065 are represented with a "strikethrough."

BEST COPY AVAILABLE

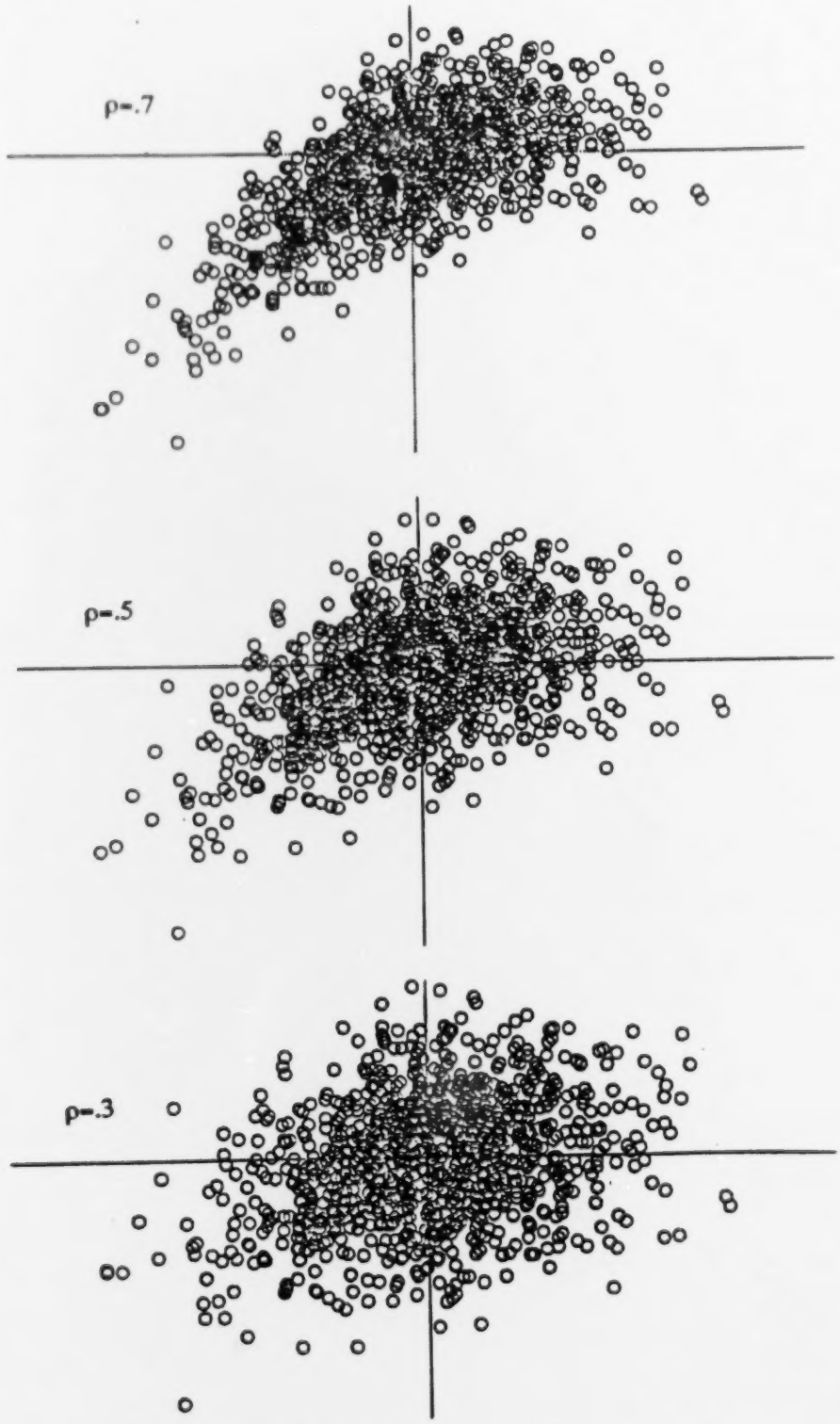
Table 3: Type I error rates and power estimates for the test of interaction

		Type I Error					Power				
Cor	b	Random	Sys.	Higher Order	RB	RB(L)	Random	Sys.	Higher Order	RB	RB(L)
.3	0	.049	.050	.050	.050	.050	.132	.139	.101	.094	.118
	2	.055	.049	.050	.049	.052	.141	.139	.10	.092	.118
	4	**.068**	.048	.050	.051	.052	~~.154~~	.133	.102	.092	.118
.5	0	.051	.049	.049	.049	.054	.371	.394	.261	.214	.299
	2	**.065**	.048	.049	.050	.055	~~.381~~	.382	.260	.210	.294
	4	**.112**	.043	.051	.052	.061	~~.406~~	.350	.263	.199	.285
.7	0	.050	.050	.051	.050	.060	.791	.818	.640	.480	.623
	2	**.090**	.045	.048	.052	.063	~~.762~~	.788	.642	.460	.604
	4	**.201**	.037	.050	.053	**.069**	~~.718~~	.698	.641	.408	~~.557~~

This table shows the Type I error rates and power estimates for the test of interactions when the correlation between x and y is .3, .5, and.7. The values of curves represents the amount of curvature present in the relationship between x and y, see figure 1 to see how much curve is represented by each value 0-4. The most heterogeneous slopes were used to determine the power. Power estimates that correspond to error rates above .065 are represented with a "strikethrough."

BEST COPY AVAILABLE

Figure 1 Scatter diagrams of maximum curvature for each value of ρ.



Note: All correlations determined prior to adding curvature.

BEST COPY AVAILABLE